

اهمیت استفاده از داده‌کاوی برای پیش‌بینی عملکرد استاد

شهرام محمدی

کارشناس ارشد کامپیوتر نرم افزار

نام نویسنده مسئول:

شهرام محمدی

تاریخ دریافت: ۱۴۰۲/۰۱/۱۶

تاریخ پذیرش: ۱۴۰۲/۰۳/۲۴

چکیده

در دهه‌های اخیر داده‌کاوی به یکی از ابزارهای مؤثر برای تحلیل داده و سیستم مدیریت دانش تبدیل شده است به طوری که بسیاری از حوزه‌ها روش داده‌کاوی را برای حل مسائلشان اتخاذ کرده‌اند. استفاده از داده‌کاوی در آموزش جهت ارتقاء سیستم آموزشی هنوز نسبتاً جدید است. این مقاله بر پیش‌بینی عملکرد استاد و بررسی عوامل مؤثر بر موفقیت دانشجویان جهت بهبود کیفیت سیستم آموزشی تمرکز دارد. مجموعه داده‌های ارزیابی دانشجویانی از ترکیه در نظر گرفته شده و کلاسبدهای داده‌ی مختلفی مانند درخت تصمیم J48، پرسپترون چند لایه، نیو بیز و بهینه‌سازی متوالی کمینه بر روی این داده‌ها اجرا شدند. مقایسه‌ای از چهار کلاسبند برای پیش‌بینی دقت و یافتن بهترین الگوریتم کلاسبند در میان آن‌ها انجام شده است. نتایج این مطالعه بسیار امیدبخش بوده و دیدگاه دیگری برای ارزیابی عملکرد دانشجویان ارائه می‌نماید. آن همچنین اهمیت به‌کارگیری ابزارهای داده‌کاوی در زمینه‌ی آموزش را برجسته می‌نماید. نتایج نشان می‌دهد که استفاده از روش ارزیابی مشخصه بر روی مجموعه داده، دقت عملکرد پیش‌بینی را افزایش می‌دهد.

واژگان کلیدی: داده‌کاوی، درخت تصمیم، پرسپترون چند لایه، نیو بیز، بهینه‌سازی متوالی کمینه.

مقدمه

امروزه داده‌کاوی (DM) توجه بسیاری را در حوزه‌ی تحلیل به خود جلب کرده و آن به ابزار جدید شناخت‌پذیری برای تحلیل داده تبدیل شده و می‌توان آن را برای استخراج دانش معنی‌دار و ارزشمند به کار برد. داده‌کاوی روش‌های امیدوارکننده‌ای برای کشف الگوهای پنهان در حجم زیادی از داده‌ها ارائه می‌نماید. این الگوهای پنهان می‌توانند به طور بالقوه‌ای برای پیش‌بینی رفتار آینده استفاده شوند [1]. بر این اساس، داده‌کاوی توسط بسیاری از محققان برای حل مسائل واقعی در حوزه‌های مختلف از قبیل بازاریابی، بازار سهام، ارتباطات راه دور، صنایع، مراقبت‌های بهداشتی، پزشکی و ارتباط با مشتری اتخاذ شده است. اخیراً تعداد معقولی از تحقیقات برای اعمال تکنیک‌های داده‌کاوی در حوزه‌ی آموزش به منظور دسته‌بندی و پیش‌بینی عملکرد دانشجویان در مؤسسات آموزشی متعدد انجام شده‌اند. به‌کارگیری تکنیک‌های داده‌کاوی در آموزش به دلیل اینکه فرصت‌های فوق‌العاده‌ای در این زمینه دارد امیدبخش است.

علاوه بر این، سیستم‌های آموزشی خواستار روش‌های جدیدی که کیفیت، کارایی و موفقیت را بهبود می‌دهند هستند [2]. عمدتاً داده‌کاوی در آموزش برای بررسی تأثیر استراتژی‌های آموزشی بر روی دانشجویان و اینکه چگونه دانشجویان درس را درک می‌کنند استفاده می‌شود [3].

عملکرد آکادمیک دانشجویان مبتنی بر عوامل مختلفی است. مهم‌ترین عوامل مشخصه‌هایی شامل رکوردهای دانشگاهی قبلی، وضعیت اقتصادی، سابقه‌ی خانوادگی، داده‌های جمعیت شناختی و روش‌های پیش‌بینی هستند. بنابراین بسیاری از تحقیقات در این زمینه مربوط به مشخصه‌های خاصی از داده‌های دانشجویان هستند. این مقاله در تلاش است اطلاعات مربوط به ارزیابی دانشجویان را برای اساتید جهت بهبود کیفیت آموزش و نشان دادن عواملی که بر عملکرد دانشجویان مؤثر است بررسی نماید. پیش‌بینی عملکرد دانشجویان به طور عمده به کیفیت روند تدریس مربوط است [4]. در این مقاله برخی از الگوریتم‌های کلاس‌بندی داده‌ها به مجموعه داده‌های ارزیابی دانشجویان ترکیه برای پیش‌بینی موفقیت دانشجویان، بررسی عملکرد استاد و پیدا کردن بهترین الگوریتم کلاس‌بندی با دقت بالا اعمال شده‌اند.

پیشینه

در سال‌های گذشته، مطالعات متعددی به بررسی داده‌کاوی برای اهداف آموزشی پرداخته‌اند. مینایی-بیدگلی یکی از اولین نویسندگانی هستند که از الگوریتم ژنتیک برای کلاس‌بندی عملکرد دانشجویان به منظور پیش‌بینی نمرات نهایی آن‌ها استفاده کردند [5]. داده‌کاوی در آموزش از سال ۱۹۹۵ تا ۲۰۰۵ در [3] که مقاله‌ی تحقیقاتی مهمی در این زمینه می‌باشد بررسی شده است. موفقیت تحصیلی دانشجویان (طبقه بندی شده به کلاس‌های پایین، متوسط و با ریسک بالا) با استفاده از روش‌های داده‌کاوی مختلف مانند درخت تصمیم (DT)، شبکه عصبی (NN) در [6] پیش‌بینی شده است. پژوهش انجام شده در [7] در تلاش برای بررسی دلیل شکست در دو کلاس هسته (ریاضی و پرتقالی) از دانش‌آموزان دو دبیرستان از منطقه‌ی آلتنجو پرتقال است. نتیجه نشان داد که هر دو الگوریتم NN و DT دقت پیش‌بینی ۷۲٪ برای یک مجموعه داده‌ی چهار کلاس دارند.

کارآمدترین روش‌های یادگیری ماشین در پیش‌بینی نمره نهایی دانشجویان کارشناسی ارشد دانشگاه انفورماتیک ایونی در [8] مورد بررسی قرار گرفت. مشخص شد که نیو بیز (NB) و الگوریتم K نزدیک‌ترین همسایه هنگامی که تعداد نمونه‌های مورد بررسی کم باشند عملکرد نمرات نهایی دانشجویان را به دقت پیش‌بینی می‌کنند. در تحقیق انجام شده در [9] تلاش شده است که تکنیک‌های داده‌کاوی با استفاده از میکروسافت و وکا بر روی مجموعه داده‌های کوچک دانشجویان اعمال شود. نتایج نشان دادند که پیش‌بینی به طور قابل ملاحظه‌ای توسط هر دو تکنولوژی موفقیت‌آمیز است. در [10] روش جدیدی برای پیش‌بینی عملکرد دانشجویان با استفاده از تفسیر داده‌کاوی معرفی شده که از تخصیص نهفته‌ی دیریکله (LDA) و ماشین بردار پشتیبان (SVM) استفاده می‌کند. نتایج بسیار امیدوارکننده بودند.

توصیف مجموعه داده

بسیاری از پژوهش‌های گذشته اغلب از معدل تجمعی (CGPA) و ارزیابی درونی مشخصه‌ها برای پیش‌بینی عملکرد دانشجویان استفاده کرده‌اند. یکی دیگر از ویژگی‌هایی که اکثراً استفاده می‌شود ارزیابی‌های خارجی و جمعیت‌شناختی دانشجویان است. تحقیقات متعددی به منظور ارتقاء سیستم‌های آموزشی با استفاده از داده‌هایی که در بالا ذکر شد انجام شده‌اند. این مقاله در تلاش برای بهبود کیفیت سیستم آموزشی با استفاده از داده‌هایی که مربوط به ارزیابی دانشجویان برای اساتیدشان است می‌باشد. این مجموعه داده از دانشگاه ایروین کالیفرنیا (UCI) مجموعه اطلاعات یادگیری ماشین جمع آوری شده و در کل شامل ۵۸۲۰ نمره‌ی ارزیابی ارائه شده توسط دانشجویان از دانشگاه قاضی در آنکارای ترکیه است. در مجموع ۲۸ سؤال از دوره‌ی خاص و ۵ مشخصه‌ی اضافی وجود دارد. Q1-Q28 همه از نوع لیکرت هستند که در آن پاسخ‌ها به صورت {۱، ۲، ۳، ۴، ۵} می‌گیرد در حالی که مقادیر نتیجه‌ای که بزرگ‌تر از ۳ هستند بسیار خوب و مقادیر نتیجه‌ای که کمتر از ۳ هستند بد در نظر گرفته می‌شوند. در حالی که سطح مقادیر حضور به صورت {۰، ۱، ۲، ۳، ۴} در نظر گرفته شده در حالی که مقادیر کمتر از ۲ ضعیف و مقادیر برابر با ۲ متوسط و مقادیر بیشتر از ۲ خوب در نظر گرفته شده‌اند. علاوه بر این، سطح دشواری مقادیر دروس به صورت {۱، ۲، ۳، ۴، ۵} در نظر گرفته شده در حالی که مقادیر کمتر از ۳ پایین، مقادیر مساوی ۳ متوسط و مقادیر بیشتر از ۳ بالا در نظر گرفته شده‌اند. اطلاعات مشخصه‌هایی که در بالا ذکر شد به دو بخش تقسیم می‌شوند: بخش اول شامل اطلاعات کلی و بخش دوم شامل پرسش‌های ارزیابی:

- inst: شناسه‌ی استاد؛ مقادیری {۱، ۲، ۳} را می‌گیرد
- کلاس: کد دوره (توصیفگر)؛ مقادیری از {۱-۱۳} را می‌گیرد
- تکرار: تعداد دفعاتی که دانشجو این درس را گرفته است؛ مقادیر {۰، ۱، ۲، ۳، ...} را می‌گیرد
- حضور: کد سطح حضور؛ مقادیری از {۰، ۱، ۲، ۳، ۴}
- دشواری: سطح دشواری درس درک شده توسط دانشجویان؛ مقادیر {۱، ۲، ۳، ۴، ۵} را می‌گیرد
- Q1-Q28 همه نوع لیکرت هستند به این معنی که مقادیر {۱، ۲، ۳، ۴، ۵} را می‌گیرند
- Q1: محتوای درس نیمسال، روش تدریس و سیستم ارزیابی که در آغاز ارائه شده‌اند.
- Q2: اهداف و موضوعات درس در ابتدای دوره به روشنی بیان شده‌اند.
- Q3: درس ارزش میزان اعتباری که به آن اختصاص داده شده را دارد.
- Q4: درس با توجه به سیلابس درسی که در اولین روز کلاس اعلام شد آموزش داده شد.
- Q5: بحث کلاسی، تکالیف و برنامه‌های کاربردی و مطالعات رضایت‌بخش بود.
- Q6: کتاب درسی و منابع درسی دیگر کافی و به‌روز بودند.
- Q7: درس امکان کار میدانی، برنامه‌های کاربردی، آزمایشگاه، مباحث و مطالعات دیگر دارد.
- Q8: آزمون‌ها، تکالیف، پروژه‌ها و امتحانات به یادگیری کمک کرده است.
- Q9: من تا حد زیادی از کلاس لذت بردم و مشتاق بودم در درس مشارکت کنم.
- Q10: انتظارات اولیه‌ی من در مورد درس در انتهای دوره یا سال برآورده شد.
- Q11: درس برای پیشرفت حرفه‌ای من مرتبط و سودمند بود.
- Q12: درس به من کمک کرد که زندگی و دنیا را با دید جدیدی بنگرم.
- Q13: دانش استاد مرتبط و به‌روز بود.
- Q14: استاد برای کلاس آماده بود.
- Q15: استاد با توجه به طرح درس اعلام شده تدریس کرد.
- Q16: استاد به درس متعهد بوده و آن قابل درک است.
- Q17: استاد به موقع وارد کلاس‌ها می‌شد.

- Q18: استاد سخنرانی/انتقال صاف و آسانی را دنبال می‌کرد.
- Q19: استاد استفاده‌ی مؤثری از ساعات کلاس می‌کرد.
- Q20: استاد درس را توضیح داده و برای کمک به دانشجویان مشتاق بود.
- Q21: استاد رویکرد مثبتی به دانشجویان نشان می‌داد.
- Q22: استاد در مورد دیدگاه‌های دانشجویان درباره‌ی درس، باز و محترمانه بود.
- Q23: استاد به مشارکت در درس تشویق می‌کرد.
- Q24: استاد تکالیف/پروژه‌های مرتبط می‌داد و دانشجویان را راهنمایی/کمک می‌کرد.
- Q25: استاد به سؤالاتی در مورد داخل و خارج از درس پاسخ می‌داد.
- Q26: سیستم ارزیابی استاد (سؤالات میان‌ترم، پایان‌ترم، پروژه‌ها و تکالیف و غیره) به طور مؤثری موضوعات درس را می‌سنجید.
- Q27: استاد راه‌حلهایی برای امتحانات ارائه می‌نمود و آن‌ها را برای دانشجویان مطرح می‌کرد.
- Q28: استاد با همه‌ی دانشجویان به شیوه‌ای مناسب و بی‌طرفانه رفتار می‌کرد.

متدولوژی و روش‌های داده‌کاوی

متدولوژی

تکنیک‌های داده‌کاوی می‌توانند به عنوان ابزاری برای توسعه و بهبود کیفیت سیستم آموزشی و همچنین ارتقاء مدیریت منابع آموزشگاه‌ها استفاده شوند. برای مثال، سؤالات جالب متعددی برای این حوزه وجود دارد که می‌توان به آن‌ها با استفاده از داده‌کاوی پاسخ داد:

- بهبود کیفیت آموزش بر اساس ارزیابی دانشجویان از دروس امکان‌پذیر است؟
 - دوره‌ها را با چه کیفیتی می‌توان به دانشجویان برای جذب بیشتر آن‌ها ارائه نمود؟
 - دلایل اصلی مؤثر بر عملکرد استاد چیست؟
 - پیش‌بینی عملکرد استاد امکان‌پذیر است؟
 - چه عواملی موفقیت دانشجویان را تحت تأثیر قرار می‌دهند؟
- پژوهش حاضر بر دو سؤال آخر متمرکز است. مدل‌سازی عملکرد استاد روش سودمندی برای هم اساتید و هم دانشجویان است زیرا آن می‌تواند به درک بهتر دستاوردهای آموزشی کمک کند.

روش‌های داده‌کاوی

تکنیک‌های داده‌کاوی برای استخراج دانش ارزشمند از حجم زیادی از داده‌ها استفاده می‌شود. تکنیک‌های داده‌کاوی مختلفی در طول دهه‌های گذشته معرفی شده‌اند. انتخاب مناسب‌ترین تکنیک برای استخراج داده‌ها، گامی است که منجر به نقشه راه صحیحی می‌شود. هدف اصلی، ساخت یک مدل کلاسیک داده‌کاوی است که ما را قادر به دسته‌بندی عواملی که بر عملکرد دانشجویان اثر می‌گذارد می‌سازد. در این مقاله، چهار تکنیک داده‌کاوی مشهور که NB، MLP، J48 DT و SMO هستند استفاده شده‌اند.

طراحی آزمایشی

برای پیاده‌سازی متوالی چهار وظیفه‌ی کلاسیک، ما از میزکار وکا استفاده کردیم. آزمایش‌ها در ۴ مرحله‌ی متوالی انجام شدند. در مرحله‌ی اول، ارزیابی مشخصه با استفاده از الگوریتم OneR برای مشخص کردن مشخصه‌هایی که بیشترین تأثیر بالقوه بر هر کلاس در مجموعه داده دارند انجام شد [11]. روش رتبه‌بندی وکا برای توجیه رتبه‌ی مشخصه‌ها با استفاده از اعتبارسنجی ۵ طرفه به کار رفته است. نتایج نشان می‌دهند که مشخصه‌های Q23، Q27، Q21، Q22، Q28، Q25 و Q24

بیشترین تأثیر را بر مجموعه داده دارند در حالی که مشخصه‌های دیگر مانند حضور، استاد، کلاس و Q1 به دلیل اینکه کمترین تأثیر را بر مجموعه داده دارند کنار گذاشته شده‌اند.

در مرحله‌ی دوم، مشخصه‌هایی که بیشترین تأثیر را دارند (مطابق با نتایج به دست آمده از مرحله‌ی قبلی، ۲۴ مشخصه‌ای که در نظر گرفته شدند) انتخاب شده و چهار تکنیک داده‌کاوی پیشنهادی بعد از حذف آخرین ۱۰ مشخصه با اثر کمتر بر روی مجموعه داده که (حضور، استاد، کلاس، Q1، دشواری، تعداد تکرارها، Q2، Q3، Q13، Q5) هستند اجرا شدند. جدول ۱ دقت پیش‌بینی در زمانی که الگوریتم ذکر شده بر روی مجموعه داده‌ها پس از انجام ارزیابی مشخصه اعمال شده است را نشان می‌دهد.

جدول ۱. دقت پیش‌بینی نتایج پس از فرایند ارزیابی مشخصه و زمانی که الگوریتم‌ها بر روی همه‌ی مجموعه داده‌ها اجرا شدند

الگوریتم	دقت عملکرد پس از پروسه‌ی ارزیابی مشخصه برای مشخصه‌هایی با بیشترین تأثیر	دقت عملکرد در زمانی که الگوریتم‌ها بر روی همه‌ی مجموعه داده‌ها برای همه‌ی مشخصه‌ها اجرا شدند
J48 DT	۸۵.۱٪	۸۴.۸٪
NB	۸۴.۳٪	۸۳.۳٪
SMO	۸۵.۸٪	۸۴.۵٪
MLP	۸۴.۶٪	۸۲.۵٪

در مرحله سه، همه‌ی مجموعه داده‌های ارزیابی دانشجویان ترکیه آزمون شده و با چهار الگوریتم کلاسیک بیان شده تحلیل شدند. مجموعه داده به دو مجموعه که ۶۶٪ برای آموزش و ۳۴٪ برای تست استفاده شده بود تقسیم شد. مدل با استفاده از مجموعه‌ی آموزش ساخته شده و با استفاده از مجموعه‌ی تست آزمون شد. مقایسه‌ای از دقت همه‌ی کلاسبندها در جدول ۱ ارائه شده است. نتایج نشان دادند که الگوریتم J48 DT به عملکرد بهتری در مقایسه با الگوریتم‌های دیگر با دقت ۸۴.۸٪ دست یافته است.

در مرحله‌ی چهار، برخی آزمایش‌ها برای بررسی عملکرد استاد انجام شدند. هدف از این آنالیز، تعیین عملکرد فردی هر استاد و بررسی عوامل مؤثر بر موفقیت آن‌ها است. چهار الگوریتم پیشنهادی بر روی مجموعه داده‌ای که به صورتی که در زیر توضیح داده شد سازمان دهی شده است اجرا شدند. اطلاعات ارزیابی از دروس که توسط هر استاد آموزش داده شده است در یک فایل مجموعه داده ترکیب شدند. از آنجایی که ما سه استاد داریم، مجموعه داده به سه فایل مجزا گروه‌بندی شد. نتایج آزمایشات به طور خلاصه در جدول ۲ و ۳ نشان داده شده‌اند.

جدول ۲. اساتید، دروس، و تعداد دانشجویانی که برای هر استاد ارزیابی شدند

استاد	کد درس	تعداد کل دانشجویان
۱	۱۰، ۷، ۲	۷۷۶
۲	۱۳، ۱۱، ۶، ۱	۱۴۴۴
۳	۱۳، ۱۲، ۹، ۸، ۵، ۴، ۳	۳۶۰۱

جدول ۳. دقت عملکرد فردی هر استاد

الگوریتم	دقت عملکرد برای استاد ۱	دقت عملکرد برای استاد ۲	دقت عملکرد برای استاد ۳
J48 DT	۸۵.۴٪	۸۵.۷٪	۸۲.۸٪
NB	۸۵.۵٪	۸۶.۸٪	۸۲.۰٪
MLP	۸۶.۲٪	۸۷.۴٪	۸۲.۸٪
SMO	۸۷.۰٪	۸۵.۴٪	۸۳.۰٪

در مرحله‌ی پنجم، چهار الگوریتم پیشنهادی بر روی مجموعه داده‌ی هر استاد به صورتی که در مرحله‌ی چهارم پس از حذف مشخصه‌هایی با بدترین رتبه که کمترین اثر بر مجموعه داده مانند مرحله‌ی ۲ داشتند اجرا کردیم. آن‌ها حضور، استاد، کلاس، Q1، دشواری، تعداد تکرارها، Q2، Q3، Q13، Q5 هستند و ۲۴ بهترین مشخصه در نظر گرفته شدند. نتایج در جدول ۴ بیان شده‌اند.

جدول ۴. دقت عملکرد اساتید برای مشخصه‌هایی که بالاترین اثر را بر روی مجموعه داده دارند

الگوریتم	دقت عملکرد برای استاد ۱	دقت عملکرد برای استاد ۲	دقت عملکرد برای استاد ۳
J48 DT	۸۵.۶٪	۸۶.۴٪	۸۳.۰٪
NB	۸۵.۹٪	۸۷.۳٪	۸۲.۸٪
MLP	۸۵.۶٪	۸۷.۸٪	۸۳.۵٪
SMO	۸۵.۲٪	۸۶.۴٪	۸۳.۸٪

نتایج و بحث و بررسی

روش‌ها و تکنیک‌های مختلفی از داده‌کاوی در این مقاله استفاده شدند. مجموعه داده تست شده و با استفاده از چهار کلاسبند داده‌ی مختلف که J48 DT، MLP، NB، SMO هستند آنالیز شد. مقایسه‌ی دقت همه الگوریتم‌ها در طول فرایند پیش‌بینی انجام شد. مشخص شد که استفاده از روش ارزیابی مشخصه بر روی مجموعه داده جهت پیش‌بینی عملکرد استاد سودمند است. مهم‌ترین مشخصه‌ها در مجموعه داده انتخاب شده سپس الگوریتم‌های ذکر شده در بالا بر روی مجموعه داده اجرا شدند. جدول ۱ مشخصه‌هایی که بیشترین تأثیر بر موفقیت دانشجویان دارد را نشان می‌دهد. این به این معنی است که این مشخصه‌ها اهمیت بیشتری در پیش‌بینی عملکرد استاد داشته و تجارب آن‌ها را توصیف می‌کند. از طرف دیگر جدول ۱ نشان می‌دهد که SMO با سطح دقت ۸۵.۸٪ بهتر از الگوریتم‌های دیگر عمل می‌کند. علاوه بر این، از جدول ۱ همچنین می‌توان مشاهده کرد که الگوریتم J48 DT در زمانی که بر همه‌ی مجموعه داده اعمال می‌شود با سطح دقت ۸۴.۸٪ عملکرد بهتری دارد.

موضوع جالب دیگری که از نتایج مشاهده می‌شود این است که عملکرد یک استاد اساساً تحت تأثیر تعداد دروسی است که آموزش می‌دهد. جدول ۴ نشان می‌دهد که همه‌ی الگوریتم‌های کلاسبندی دقت پیش‌بینی کمتری در زمانی که بر روی فایل مجموعه داده‌ی استاد ۳ اجرا شده‌اند در مقایسه با دقت پیش‌بینی به دست آمده از الگوریتم‌هایی که بر روی فایل مجموعه داده‌ی استاد ۱ و ۲ اجرا شده‌اند دارد. با مقایسه‌ی همه‌ی کلاسبندها، الگوریتم‌های SMO و MLP به ترتیب با دقت‌های ۸۷.۰٪ و ۸۶.۲٪ بهترین عملکرد در میان همه‌ی کلاسبندها را در مجموعه داده‌ی استاد ۱ همان‌طور که جدول ۳ نشان می‌دهد داشتند. در حالی که دقت SMO کاهش می‌یابد، MLP تا رسیدن به بهترین عملکرد با دقت ۸۷.۲٪ برای مجموعه داده‌ی استاد ۲ همان‌طور که در جدول ۳ نشان داده شده است ادامه می‌یابد. از طرف دیگر نتایج نشان می‌دهند که دقت عملکرد در زمانی که مشخصه‌های با رتبه‌ی بد حذف شده‌اند نسبت به وقتی که الگوریتم‌ها روی همه‌ی مشخصه‌های مجموعه داده اجرا شده‌اند افزایش می‌یابد. جدول ۴ دقت عملکرد اساتید برای مشخصه‌هایی که بالاترین تأثیر بر مجموعه داده دارند را پس از حذف مشخصه‌های بد نشان می‌دهد. می‌توان از نتایج استنباط کرد که دقت عملکرد همه‌ی الگوریتم‌ها در جدول ۴ بهتر از دقت به دست آمده از این الگوریتم‌ها با همه‌ی مشخصه‌هایی که در جدول ۳ است به غیر از MLP و SMO که عملکرد خوبی بر روی مجموعه داده‌ی استاد ۱ در جدول ۳ دارند در حالی که عملکرد آن‌ها در زمان اجرا بر روی مجموعه داده‌ی استاد ۱ پس از حذف مشخصه‌های بد همان‌طور که در جدول ۴ نشان داده شده است کاهش می‌یابد.

نتیجه‌گیری

ما نتیجه گرفتیم که استفاده از داده‌های ارزیابی دانشجویان برای دروس جهت پیش‌بینی عواملی که بر موفقیت آن‌ها و همچنین پیش‌بینی عملکرد اساتید مؤثر است سودمند می‌باشد. علاوه بر این، آن نقطه نظر دیگری برای بهبود کیفیت آموزشی است، که برای جذب دانشجویان حیاتی می‌باشد در حالی که اکثر محققان از معدل تجمعی و مشخصه‌های ارزیابی داخلی برای پیش‌بینی عملکرد دانشجویان جهت ارتقا سیستم آموزشی استفاده می‌نمایند. علاوه بر این حذف مشخصه‌های بد که اثر کمتری بر مجموعه داده دارند دقت عملکرد الگوریتم‌ها را افزایش می‌دهد.

منابع و مراجع

- [1] Mendes RR, de Voznika FB, Freitas AA, Novella JC. Discovering fuzzy classification rules with genetic programming and co-evolution. *Principles of Data Mining and Knowledge Discovery*, Springer Berlin Heidelberg, 2001, p. 314-325.
- [2] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 2014, vol. 41, no. 4. p. 432-462.
- [3] Andonie R. Extreme data mining: Inference from small dataset. *International Journal of Computers Communications and Control*, 2010, vol. 5, no. 3. p. 280-291.
- [4] Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 2007, vol. 33, no. 1, p. 135-146.
- [5] Minaei-Bidgoli B, Punch WF. Using genetic algorithms for data mining optimization in an educational web based system. *Genetic and Evolutionary Computation*, Springer Berlin Heidelberg, 2003, p. 2252-2263.
- [6] Superby JF, Vandamme JP, Meskens N. Determination of factors influencing the achievement of the first-year university students using data mining method. *Workshop on Educational Data Mining*, 2006, p. 37-44.
- [7] Cortez P, Silva A. Using data mining to predict secondary school student performance. *EUROSTAT, A. Brito and J. Teixeira (Eds.) 2008*, pp. 5-12.
- [8] Koutina M, Kermanidis K L. Predicting postgraduate students' performance using machine learning technique. *Artificial Intelligence Applications and Innovations*, Springer Berlin Heidelberg 2011, p. 159-168.
- [9] Natek S, Zwillig M. Student data mining solution—knowledge management system related to higher education institution. *Expert systems with applications*, 2014, vol. 41, no.14, p. 6400-6407.
- [10] Sorour SE, Mine T, Goda K, Hirokawa S. Estimation of student performance by considering consecutive lesson. *4th International Congress on Advanced Applied Informatics*, 2015, p.121-126.
- [11] Forman G, Cohen I. Learning from little: Comparison of classifiers given little training. *Knowledge Discovery in Databases: PKDD*, Springer Berlin Heidelberg, 2004, p. 161-172.